02. Review of Statistical tests and ANOVA

Quantitative Genetics Exercise, Universität Potsdam, Dr. Hao Tong

1. Fill up the below ANOVA table and answer questions.

Source	Sum of Square	Degree of freedom	Mean Square	F-score
Treatment		3		
Error			30	
Total	1200	19		

How many treatment groups? How many samples in each group if the data is balanced and without missing?

What is the corresponding p-value for the F-score in the table? Is the difference of means between treatments significant at level α=0.05?

2. Read the data set of maize yield (Ex2.csv) and format it to the data frame that ANOVA required. For this question, **please use the first three columns in raw data corresponding to three environments measurement in first family only**. Assume that all samples are independent, please write the one-way ANOVA table for environment factors. What conclusion you can make from this ANOVA table?

3. Use the data from previous exercise (Ex1.csv) to build regression model for body weight against body height, as below:

$$Weight = X + Height \times \beta + \varepsilon.$$

(1) What is the value of t-statistic of regression coefficient of intercept and height, respectively? How did these values be calculated? Write the corresponding numerical equations based on the regression summary table.

(2) Try to write the ANOVA table for this regression. What is the value of degree of freedom of each coefficient, and why? Explain in numerical equations. What is the value of F-statistic? Is it significant at level α=0.05? Try to calculate the Coefficient of determination (R^2) based on the ANOVA table.

4. Are the mean values in three environment groups significantly different between each other? Try to use several multiple comparison methods to clarify,

including Bonferroni, Holm, HSD. Try to understand the HSD method, what is the principal idea of HSD method? Are the conclusions from different methods identical?

5. For this question, **please use all raw data corresponding to two factors: environment and family**. Assume that all samples are independent, please write the two-way ANOVA tables with or without interaction, respectively in R. Please write the equations how to calculate the degree of freedom of each factor, and try to understand how to get the sum square of each factor in the two-way with interaction ANOVA table. What conclusions you can make from the ANOVA tables with or without interaction, respectively?